**Introduction**

In this report the performance of the proposed K-Nearest Neighbors (KNN) classifier is assessed using a simulated dataset. Our goal here was to determine how the accuracy of the classifier will improve as the values of k for the number of neighbors changes. This dataset was created with the help of the make blobs function, placing it into three distinct clusters with centers at (3, 3), (7, 7) and (10, 2); in total there are 309 samples in the experiment.

**Procedure**

**1. Preprocessing and data generation:**

* The dataset was generated using make blobs function implemented in the sklearn. datasets.
* Dataset characteristics:
* It is a dataset with a total of 309 samples whereby all of them fall under the three cluster categories.

The centers of the clusters were chosen to be (3,3); (7,7) and (10,2).

**2. Training and Testing Split:**

* The data was split using the train\_test\_split function from sklearn. The features and labels of the data was extracted for modelling. splitting the dataset into training and testing sets, using model selection where 80% of the data would be used for training while the remaining 20% would be used for testing. This makes it possible to later test the model on new data which it hasn’t seen before.

**3. K-Nearest Neighbors Classifier Implementation:**

* Classifier Initialization:

Implementation of the K-Nearest Neighbors (KNN) classifier was done with the help of sklearn package. neighbors. KNeighborsClassifier. As we defined in the first model, the initial number of neighbors, k, is taken to be 8.

o Model Training:

The classifier was trained to the training dataset making use of the fit() technique.

o Prediction:

The labels for the test set were predicted using the ‘predict()’ method.

o Accuracy Evaluation:

The accuracy score from sklearn. as the evaluation criterion, there was used ‘metrics’ that describes the quantifiable measure of how well the predicted labels matched the test set labels.

**4. Exploring Accuracy for Different k Values:Exploring Accuracy for Different k Values:**

o We analysed the result of the classifier for varying the number of neighbours k, from 1 to 10. As for every specific k, the accuracy of the classifier was calculated, and the line plot of the results obtained was drawn. It is shown in this plot how the number of neighbours impacts the classification rate according to this plot.

**Results**

1. Accuracy of the KNN Classifier (k=8):Accuracy of the KNN Classifier (k=8):

The KNN classifier with parameter setting of k=8 was computed to achieve an accuracy of 0. 97 meaning that the model accurately classified the test data.

2. Effect of k on Accuracy:Effect of k on Accuracy:

o The efficiency of the classifier was measured for the values of k between 1 and 10.

o Key Observations:

The measures of accuracy are consistently high over different values of ‘k’ thus the stability of the method in its chosen value for k is upheld.

As it can be seen in the table below, when k=8 and k=7 the classifier worked best but as k grew slightly, minor changes were noted.

o Plot:

To display the accuracy of each k dominant, a line plot was made as this would give an indication of the type of neighbors affects the performance of the data set. From the graph it is evident that the horizontal speed of the model is reasonably steady and has not dropped from the efficiency of its classification.

A graph with a line

Description automatically generated

Conclusion

From this experiment, the efficiency of applying the KNN classifier on the synthetic data with distinguishable clusters is shown. The classifier showed high accuracy of results during subsequent iterations, and the first created model with the number of clusters k=8 yielded 95% of accuracy. This it is apparent the simplicity of the dataset enabled the classifier to exhibit acceptable behavior for all the varieties of k that were tried, and as such the choice of the right k does not appear to be a huge issue this particular circumstance. Other experiments with more intricate and intertwined data sets as to the current study might reveal a better understanding of KNN performance with more complicated classifications.